

AD-A065 459

TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES
M.D. I. ESTIMATION VIA UNCONSTRAINED CONVEX PROGRAMMING.(U)
NOV 78 P L BROCKETT, A CHARNES, W W COOPER

F/G 9/4

N00014-75-C-0616

UNCLASSIFIED

CCS-326

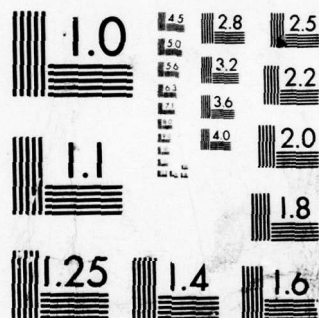
NL

| OF |
AD
A065 459



END
DATE
FILMED

4 --79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

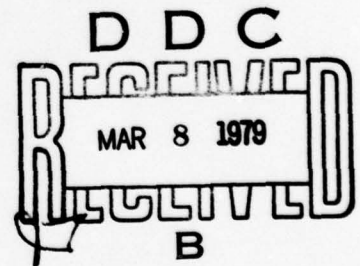
LEVEL II

AD A0 65459

DDC FILE COPY

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712



DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited



79 03 01 010

(12) LEVEL II

DDC FILE COPY ADA065459

Research Report CCS 326

M. D. I. ESTIMATION VIA
UNCONSTRAINED CONVEX PROGRAMMING

by

P. L. Brockett *

A. Charnes *

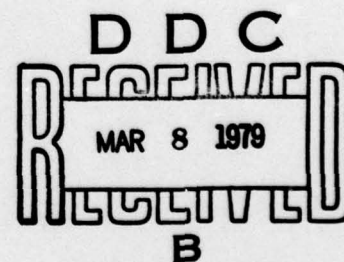
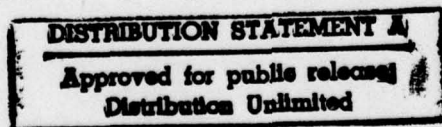
W. W. Cooper **

November 1978

*University of Texas at Austin

**Graduate School of Business Administration, Harvard University, Boston, Mass.

This research was partly supported by Project NR047-021, ONR Contracts N00014-75-C-0616 and N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.



CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

79 03 01 010

Abstract

A method is presented for obtaining minimum discrimination information (M.D.I.) estimates of probability distributions. This involves using an extremal principle of Charnes and Cooper [4] and, viewing M.D.I. estimation in a dual convex programming framework. The resulting dual convex program is unconstrained and involves only exponential and linear terms, and hence is easily solved. This approach makes M.D.I. estimation computationally efficient and reduces the time and cost of obtaining such estimates.

Key words and phrases: minimum discrimination information estimation, information theory, convex programming, unconstrained dual

ACCESSION for		
NTIS	White Section	<input checked="checked" type="checkbox"/>
DDC	Buff Section	<input type="checkbox"/>
UNANNOUNCED		<input type="checkbox"/>
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AUTH. or	SPECIAL
A		

Wiener (1948 p. 76) remarked quite early that entropy (or Shannon-Wiener type measures of the amount of information) could eventually replace Fisher's definition [9] of information (see Kullback [11]). Still, information theory is often mistakenly considered primarily as a subfield of communication theory, where indeed Shannon's entropy has proved essential [14]. The statistical community has only fairly recently (after the 1967 translation of Kullback's 1959 monograph into Russian) been assessing the use of information theoretic concepts in inference, although there were notable early recognitions of the statistical power of the theory (e.g. [12], [13], [16], see also the references in [11]). The information functional we consider will be called the Khinchin-Kullback-Leibler functional in honor of their early contribution to this theory. Modern contributions of Akaike elucidate some of this power, and show that the information theoretic framework is perhaps the proper approach to many diverse problems of statistics. In [1] he gives an information theoretic extension of the maximum likelihood principle and shows that the Khinchin-Kullback-Leibler type information functional naturally arises in statistical problems. By utilizing the maximum relative entropy (or equivalently the minimum expected log likelihood ratio) quantity he is able to encompass both statistical estimation and hypothesis testing into a decision theoretic

framework using a Khinchin-Kullback-Leibler type loss function. His techniques are applied to such important considerations as the decision of the number of factors to include in factor analysis, the number of independent variates to choose in multiple regression, and the order of the model when fitting an autoregressive time series. In [2] Akaike shows that using his extension of the maximum likelihood method enables one to obtain a solution to the problem of James-Stein estimators. In [3] he looks at Bayes procedures from an information theoretic point of view.

The information theoretic approach is based upon the mean information for discriminating between two densities f_1 and f_2 (relative to some fixed dominating measure λ). The mean information for discriminating in favor of f_1 against f_2 is defined by Kullback [11] as

$$I(f_1 : f_2) = \int f_1(x) \ln \left[\frac{f_1(x)}{f_2(x)} \right] \lambda(dx).$$

He also calls this the directed divergence between the two probability measures and shows $I(f_1 : f_2) \geq 0$ with equality if and only if $f_1 = f_2$ a.e. $[\lambda]$. Thus one may estimate f_2 by that density f_1 which is closest in the sense of information distance to f_2 , and one may impose additional constraints upon f_1 when necessary. This method of estimation is called minimum discrimination information (M.D.I.) and is based upon the following inequality:

M.D.I. inequality: Suppose $T(x)$ is a statistic for which $M_2(\tau) = \int e^{\tau T(x)} f_2(x) \lambda(dx)$ exists in an interval, and consider those

densities f_1 satisfying $\theta = \int T(x) f_1(x) \lambda(dx)$ for a known parameter θ .

Then $I(f_1:f_2) \geq \theta \tau - \ln M_2(\tau)$ where $\theta = \frac{d}{d\tau} \ln M_2(\tau)$ with equality iff

$$f_1(x) = f_1^*(x) = e^{\tau T(x)} f_2(x) / M_2(\tau) \quad \text{a.e. } [\lambda].$$

The density $f_1^*(x)$ is called the M.D.I. estimate of f_2 subject to $\theta = \int T(x) f_1(x) d\lambda(x)$ (or the "conjugate" distribution in Khinchin's terminology).

Since solving for the M.D.I. estimate f_1^* entails solving the equation $\frac{d}{d\tau} \ln M_2(\tau) = \theta$ for τ (a highly non linear problem) it has been quite difficult computationally to obtain M.D.I. estimates. This implicit differential equation relation is also difficult to work with analytically.

The purpose of this paper is to show how to view M.D.I. estimation from a dual convex programming point of view and to point out analytical properties of the estimates which follow directly from the form of this duality. In particular, the dual problem is unconstrained and involves only exponential and linear terms. This pair of dual problems is easily solved by any of a number of existing algorithms. The τ variables needed for f_1^* and called "dual parameters" in Gokhale and Kullback (1978) are here shown to be actual variables in the dual convex programming problem.

This dual formulation was first developed by Charnes and Cooper [4], [5] (see also [7]) who applied the technique to show that the accounting balance equations for a cartel or "resource-value transfer" economy

could be derived from Khinchin-Kullback-Leibler statistical estimates constrained by a linear inequality system. Other new developments showing that old heuristic estimating procedures are actually constrained Khinchin-Kullback-Leibler estimations, such as (Charnes, Raike, Bettinger [8]) "gravity potential" estimates, SANDDABS estimates in marketing analysis (Charnes, Cooper, Learner [6]), and various depreciation methods in accounting (Theil and Lev [15]) lend great weight to the idea that efficient analytical and computational techniques for M.D.I. estimation can be valuable in applied research.

An important application of M.D.I. estimation is to the analysis of contingency tables, and we shall utilize this example to elucidate the techniques of this section. Denote a contingency table cell by the generic label w and the collection of all cells by Ω . For a suitable choice of probability measure $\pi(w)$ over the contingency table $w \in \Omega$ (in general $\pi(w)$ is determined by the specific problem of interest), Gokhale and Kullback [10] pose the problem of finding that probability measure p^* (the M.D.I. estimate) which minimizes $I(p; \pi)$ subject to the equality constraints $\mathcal{C} p = \underline{\theta}$ and the non-negativity constraints $p \geq 0$. Here \mathcal{C} (called the design matrix) is an $(r + 1) \times |\Omega|$ matrix, p is the $1 \times |\Omega|$ vector of probabilities and $\underline{\theta}$ is a $1 \times (r + 1)$ vector. If we denote the rows of \mathcal{C} by $C_i(w)$, $i = 0, \dots, r$, the constraints are of the form $\sum_{\Omega} C_i(w) p(w) = \theta_i$, $i = 0, \dots, r$. Usually we take $\theta_0 = 1$ and $C_0(w) = 1$ for all $w \in \Omega$ so that the first constraint insures that p is a probability measure. The remaining r equations are (usually) moment constraints. Gokhale and Kullback additionally require that the vectors $C_i(\cdot)$ are linearly independent in order to obtain their estimates. As we shall see, the programming formulation does not require this assumption. This is important since, even aside from having to recognize linearly independent constraints, the transformation to obtain such linearly independent constraints can be difficult and onerous.

By using Lagrange multipliers Gokhale and Kullback show that the minimizing probability distribution is

$$(2.1) \quad p^*(\omega) = \exp \{ \tau_0 + \tau_1 C_1(\omega) + \dots + \tau_r C_r(\omega) \} \pi(\omega)$$

where, they say, the τ 's are to be determined in such a way that $\underline{C}p^* = \underline{\theta}$ holds.

The problem of determining (τ_0, \dots, τ_r) so as to satisfy $\underline{C}p^* = \underline{\theta}$ is in general, quite difficult. The following Charnes-Cooper extremal principle, however, makes this very easy computationally.

Let $K(\delta, x) = de^x - \delta x$ for $\delta \geq 0$, $d > 0$, $x \in \mathbb{R}$, and define $g(\delta) = \inf_x K(\delta, x) = \delta - \delta \ln \left(\frac{\delta}{d} \right) = -\delta \ln \left(\frac{\delta}{ed} \right)$. Then it follows that for $\underline{\delta} = (\delta_1, \dots, \delta_n)'$, and $\underline{x} = (x_1, \dots, x_n)'$

$$(2.2) \quad v(\underline{\delta}) \equiv \sum_i g(\delta_i) = -\sum_i \delta_i \ln \frac{\delta_i}{ed_i} \leq \sum_i [d_i e^{x_i} - \delta_i x_i].$$

Suppose that $\underline{x} = \underline{A} \underline{z}$ for some matrix \underline{A} . And let \underline{A}_i denote the i th row of \underline{A} , and set $\underline{A}' \underline{\delta} = \underline{b}$. Then (2.2) becomes

$$(2.3) \quad v(\underline{\delta}) = -\sum_i \delta_i \ln \left(\frac{\delta_i}{ed_i} \right) \leq \sum_i [d_i e^{\underline{A}_i \underline{z}} - (\delta_i) \underline{A}_i \underline{z}] \\ = \sum_i d_i e^{\underline{A}_i \underline{z}} - \underline{b}' \underline{z} \equiv \xi(\underline{z})$$

which holds for all \underline{z} and $\underline{\delta} \in \{ \underline{\delta} : \underline{A}' \underline{\delta} = \underline{b}, \underline{\delta} \geq 0 \}$.

We then have:

Theorem 2.1 (Charnes - Cooper [4], [5]) For the following dual convex programs

$$(I) \quad \sup v(\underline{\delta}) \equiv -\sum_i \delta_i \ln \left(\frac{\delta_i}{ed_i} \right) \quad \text{subject to}$$

$$\underline{A}' \underline{\delta} = \underline{b} \\ \underline{\delta} \geq 0$$

and

$$(II) \quad \inf_{\underline{z}} \xi(\underline{z}) \equiv \sum_i d_i e^{iA\underline{z}} - \underline{b}'\underline{z}$$

there are exactly three mutually exclusive and collectively exhaustive duality states:

(1) $\Delta \equiv \{ \underline{\delta} : A'\underline{\delta} = \underline{b}, \underline{\delta} \geq 0 \} = \emptyset$ and $\xi(\underline{z})$ is unbounded below.

(2) $\Pi \delta_i = 0$ for all $\underline{\delta} \in \Delta \neq \emptyset$ and $\xi(\underline{z})$ has only an infimum.

Further $\inf \xi(\underline{z}) = \min \xi_D(\underline{z})$ where $\xi_D(\underline{z})$ contains only those terms of $\xi(\underline{z})$ for which $\delta_i > 0$ in some $\underline{\delta} \in \Delta$.

(3) There exists $\underline{\delta} \in \Delta$ with $\underline{\delta} > 0$ and $\xi(\underline{z})$ has a minimum at \underline{z}^* .

Further: (a) $v(\underline{\delta})$ has a unique maximum at $\underline{\delta}^* > 0$

(b) $v(\underline{\delta}^*) = \xi(\underline{z}^*)$

(c) $\delta_i^* = d_i e^{iA\underline{z}^*}$

Note that there is no requirement of linear independence made, and all possible behaviors of the system Δ are comprehended. Of course, the usual state in applications will be (3). If the requisites for state determination are not obvious, the state may be determined by solution of the linear programming problem: $\max \mu$

subject to $\underline{\mu} \underline{e}' - \underline{\delta}' \leq 0$

$A'\underline{\delta} = \underline{b}$

$\underline{\delta} \geq 0$

State (1) corresponds to infeasibility, state (2) corresponds to $\mu^* = 0$ and state (3) corresponds to $\mu^* > 0$. (c.f. [7])

This result is very attractive since the dual problem (II) is an unconstrained convex programming problem involving only exponential and linear terms and hence is easily solved numerically. The original constrained Khinchin-Kullback-Leibler estimation problem (I) was very difficult to solve. It should also be noted that case 3 of Theorem 2.1 could equally well have been proven in the general measure-theoretic case in almost the same manner as the finite discrete case [7], [17].

Referring back to (2.1) and the determination of (τ_0, \dots, τ_r) , we note that taking $\delta_w = p(w)$, $w \in \Omega$, $d_w = e^{-1} \pi(w)$, $\tilde{A} = \tilde{C}$ and $\tilde{b} = \tilde{\theta}$ in Theorem 2.1 yields the Gokhale-Kullback M.D.I. problem mentioned at the beginning of this section. The dual problem is the unconstrained

$$(2.4) \quad \min_{\tilde{z}} \sum_w e^{-1} \pi(w) e^{\tilde{C}_w \tilde{z}} - \tilde{\theta}' \tilde{z}$$

where \tilde{C}_w is the w th column of the $(r+1) \times \Omega$ matrix \tilde{C} . If

this minimum occurs at \tilde{z}^* say, then theorem 2.1 implies that

$$\tilde{z}^* \text{ is to be identified as } \tilde{\tau}^* \text{ in } p(w) = e^{-1} \pi(w) e^{\tilde{C}_w \tilde{\tau}^*} = \\ \exp \{ \tau_0^* + \tau_1^* C_1(w) + \dots + \tau_r^* C_r(w) \} \pi(w).$$

Thus the τ 's arise in (2.1) and are easily computed from the unconstrained convex programming problem (2.4). Thus the computation of constrained M.D.I. estimates becomes a rather straightforward problem.

Further, item (c) in state 3 shows immediately that whatever the form of the d_i - distribution, the M.D.I. estimated distribution is of the exponential family. As may be noted from the results for general

distributions in [7], a similar property persists there as well. Thus the M.D.I. estimated distributions have the attractive property of preserving sufficient statistics associated with the target distribution since its density is multiplied by an exponential linear in \tilde{z}^* to get the M.D.I. estimate.

Also, Gokhale and Kullback [10] noted, when fitting marginals for contingency tables by M.D.I. estimation one obtains the log-linear model for contingency table entries as a by-product. Here this is immediately evident (without further qualifications or Gokhale and Kullback's additional requirements) by taking logs in item (c) of state 3.

As alluded to in section one, this dual approach to constrained Khinchin-Kullback-Leibler estimation has already proven valuable in practice. In marketing research, Charnes, Cooper and Learner [6] substantially extended the SANDDABS analysis used to evaluate consumer purchase behavior and brand shifting. They showed that the calculations involved in the analysis could be viewed as constrained Khinchin-Kullback-Leibler estimation, and thereby brought the procedure under the ambit of statistical theory rather than leaving it as merely a heuristic estimation method. By utilizing the dual approach outlined in Theorem 2.1 they were able to improve the computational efficiency of the analysis as well.

REFERENCES

- [1] Akaike, H. "Information Theory and an Extension of the Maximum Likelihood Principle," in Proc. of the 2nd Int. Sym. on Info. Theory [1972].
- [2] Akaike, H. "An Extension of the Method of Maximum Likelihood and the Stein's Problem," Ann. Inst. Statist. Math 29 (1977) part A, 153-164.
- [3] Akaike, H. "A New Look at the Bayes Procedure," Biometrika 65 (1978) 1, 53-59.
- [4] Charnes, A. and W. W. Cooper. "An Extremal Principle for Accounting Balance of a Resource Value Transfer Economy: Existence, Uniqueness and Computation," Accademia Nazionale dei Lincei, Rome, Series VIII, Vol. LVI 1974, 556-561.
- [5] Charnes, A. and W. W. Cooper. "Constrained Kullback-Leibler Estimation; Generalized Coob-Douglas Balance and Unconstrained Convex Programming," Accademia Nazionale dei Lincei, Rome, Series VIII, Vol. LVIII 1975, 568-576.
- [6] Charnes, A., W. W. Cooper, and D. B. Learner. "Constrained Information Theoretic Characterizations in Consumer Purchase Behavior," Journal of the Operational Research Society, 1978.
- [7] Charnes, A., W.W. Cooper and L. Seiford. "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leibler Estimation," Math. Operationsforsch. Statist. Ser. Optimization, Vol. 9 (1978) No. 1, 21-29.
- [8] Charnes, A. and W.M. Raike and C.O. Bettinger. "An Extremal and Information Theoretic Characterization of Some Interzonal Transfers," Socio-Economic Planning Sciences, Vol. 6, 1972, 531-537.
- [9] Fisher, R. A. "Theory of Statistical Estimation," Proc. Camb. Phil. Soc., Vol. 22, 700-725; Contributions to Mathematical Statistics, New York: John Wiley & Sons, 1950, paper 11.
- [10] Gokhale, D.V. and S. Kullback. The Information in Contingency Tables, New York: Marcel Dekker, Inc., 1978.

- [11] Kullback, S. Information Theory and Statistics, New York: John Wiley & Sons, 1959.
- [12] Kullback, S. and R.A. Leibler "On Information and Sufficiency," Ann. Math. Stat., Vol. 22, 1951, 79-86.
- [13] Khinchin, A.I. Mathematical Foundations of Statistical Mechanics New York: Dover Publishers,
- [14] Pierce, J.R. Electrons, Waves and Messages. New York: Hanover House, 1956.
- [15] Lev, B. and H. Theil "A Maximum Entropy Approach to the Choice of Asset Depreciation," Report 7741, July 1977, Graduate School of Business, University of Chicago.
- [16] Wiener, N. Cybernetics, New York: John Wiley & Sons, 1948.
- [17] Ben-Tal, A. and Charnes, A., "A Dual Optimization Framework for Some Problems of Information Theory and Statistics," forthcoming Journal of Problems of Control and Information Theory.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author)

The Center for Cybernetic Studies
The University of Texas at Austin

29. REPORT SECURITY CLASSIFICATION

Unclassified

29. GROUP

3. REPORT TITLE

⑥ M.D.I. Estimation Via Unconstrained Convex Programming

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

⑩ Patrick L. Brockett, Abraham Charnes, William W. Cooper

6. REPORT DATE

⑪ November 1978

7a. TOTAL NO. OF PAGES

13

7b. NO. OF REFS

17

8. CONTRACT OR GRANT NO.

N00014-75-C-0616 & 0569

9. PROJECT NO.

NR047-021

30. ORIGINATOR'S REPORT NUMBER(S)

⑨ Center for Cybernetic Studies
Research Report Number 326

31. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

⑫ 46 P1

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

⑭ 225-326

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research (Code 434)
Washington, D.C.

13. ABSTRACT

→ A method is presented for obtaining minimum discrimination information (M.D.I.) estimates of probability distributions. This involves using an extremal principal of Charnes and Cooper [4], and, viewing M.D.I. estimation in a dual convex programming framework. The resulting dual convex program is unconstrained and involves only exponential and linear terms, and hence is easily solved. This approach makes M.D.I. estimation computationally efficient and reduces the time and cost of obtaining such estimates. ←

⑮ N00014-75-C-0616, N00014-75-C-0569

406 297

self

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
<p>Minimum Discrimination Information Estimation</p> <p>Information Theory</p> <p>Convex Programming</p> <p>Unconstrained Dual</p>						